



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



Metadata-Centric Discovery Service

Thomas Huang, Edward Armstrong, Nga Chung

PO.DAAC/JPL



National Aeronautics and
Space Administration
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



What is This?





National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



Some Metadata



File type: JPEG image

File size: 2.7 MB (2,664,608 bytes)

Compression: None

Image size: 2592 × 1944 pixels

Creation date: Jun 13, 2008 4:57 PM

Modification date: Jun 13, 2008 4:57 PM



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



More Metadata



File type: JPEG image

File size: 2.7 MB (2,664,608 bytes)

Compression: None

Image size: 2592 × 1944 pixels

Creation date: Jun 13, 2008 4:57 PM

Modification date: Jun 13, 2008 4:57 PM

Long Name: Réserve Naturelle des Sept Îles, France

Location: 48° 53' 12" N 3° 28' 53" W (carte)

Description: Nature Reserve in the archipelago of seven islands, the largest bird sanctuary in France

Additional information: http://fr.wikipedia.org/wiki/Réserve_naturelle_des_Sept-Îles





Our Challenges

- OpenSearch for now, but ...
- Our users speak different (metadata) languages
 - FGDC
 - GCMD
 - ECHO
 - ISO 19115-2
 - and many more...
- We host a lot of metadata, including some internal services for various context
- But the Internet is THE metadata repository
- A metadata architecture must be extensible to support multi metadata dialect
- A discovery architecture must be extensible to support many repositories

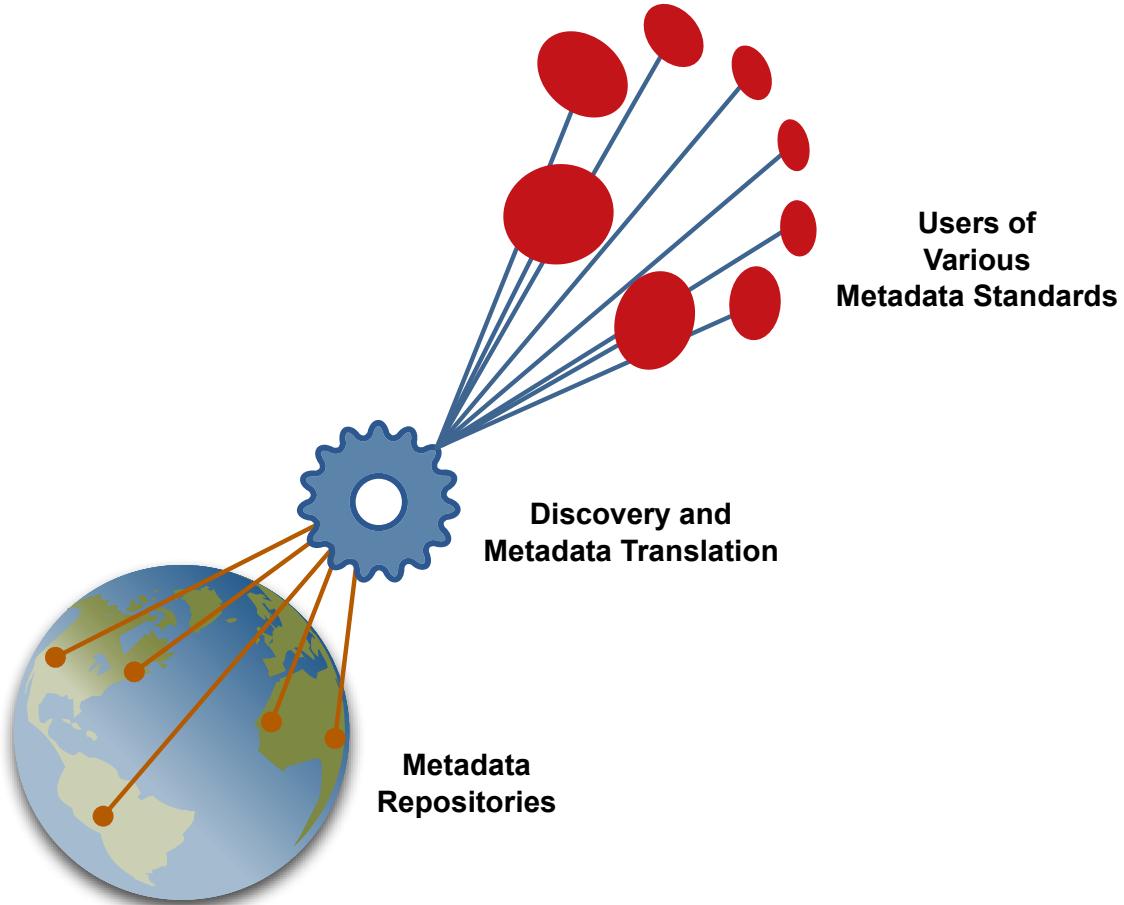


National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



A Many-to-Many World





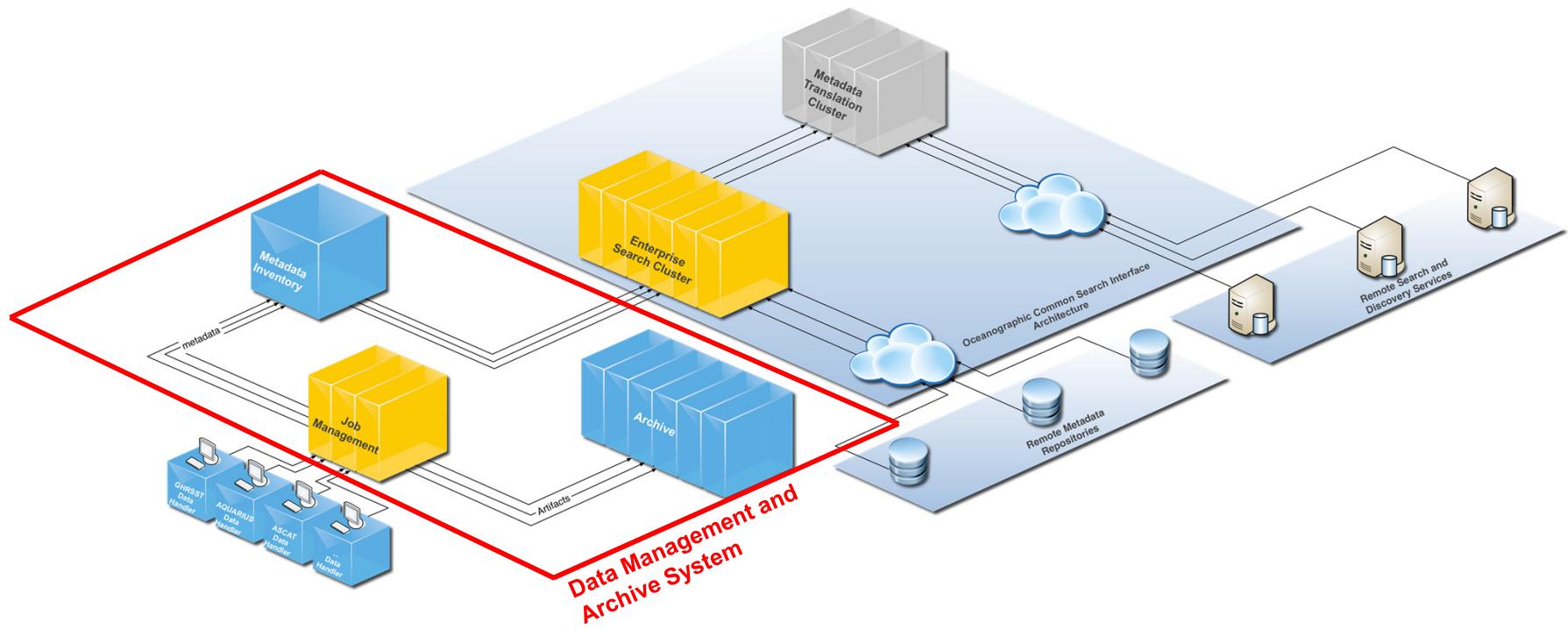
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



OCSI System Architecture

Oceanographic Common Search Interface



Aggregations and Translation



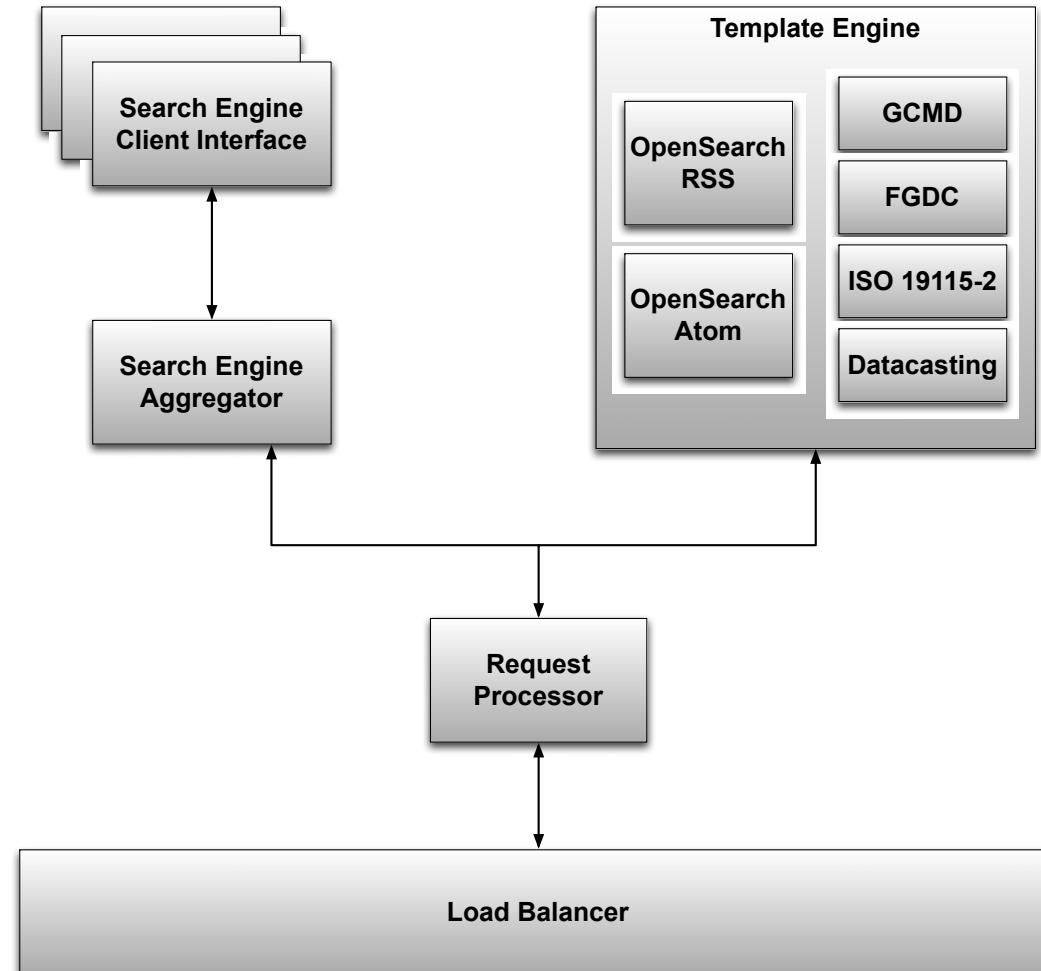
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

OCSI



METADATA TRANSLATION ARCHITECTURE





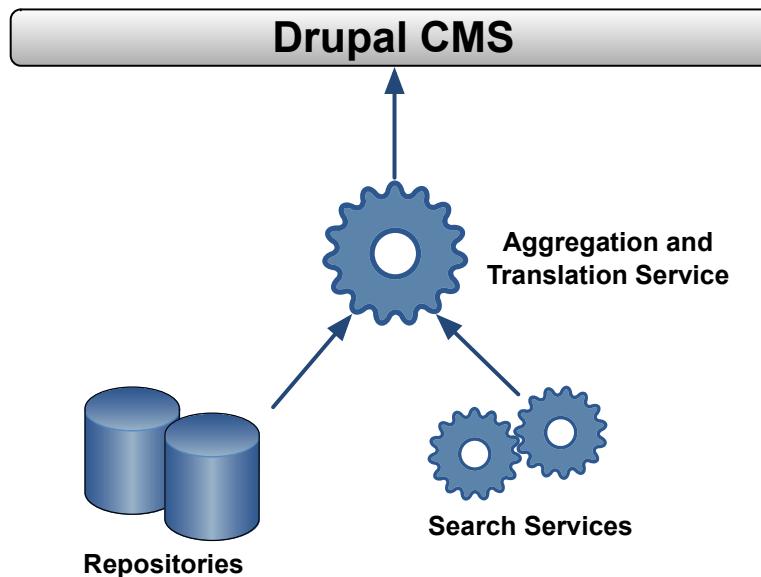
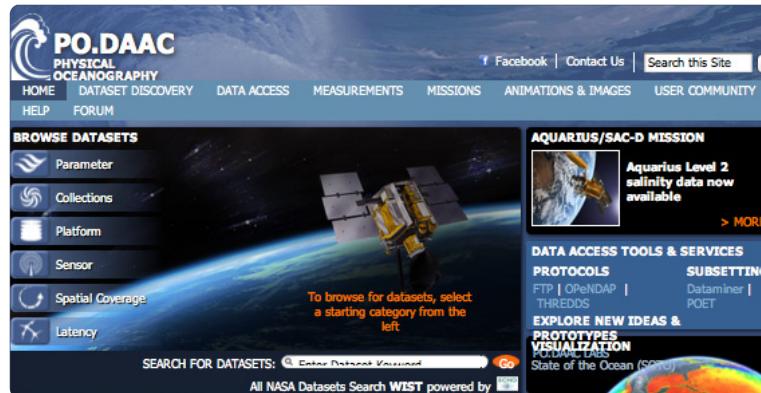
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



The PO.DAAC Web Portal

<http://podaac.jpl.nasa.gov>





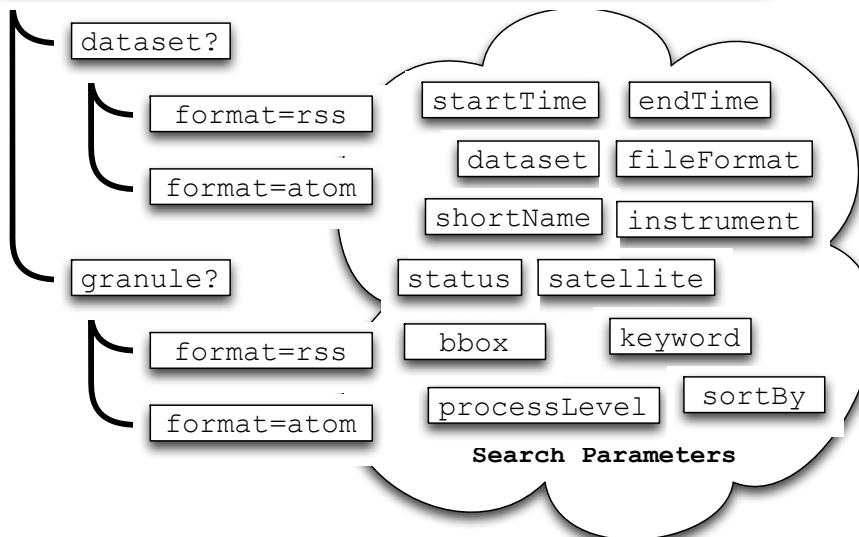
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



OpenSearch

`http://podaac.jpl.nasa.gov:8890/ws/search/`



Terminal

```
% curl -X GET \
? "http://podaac.jpl.nasa.gov:8890/ws/search/dataset?format=rss&keyword=ocean"
```

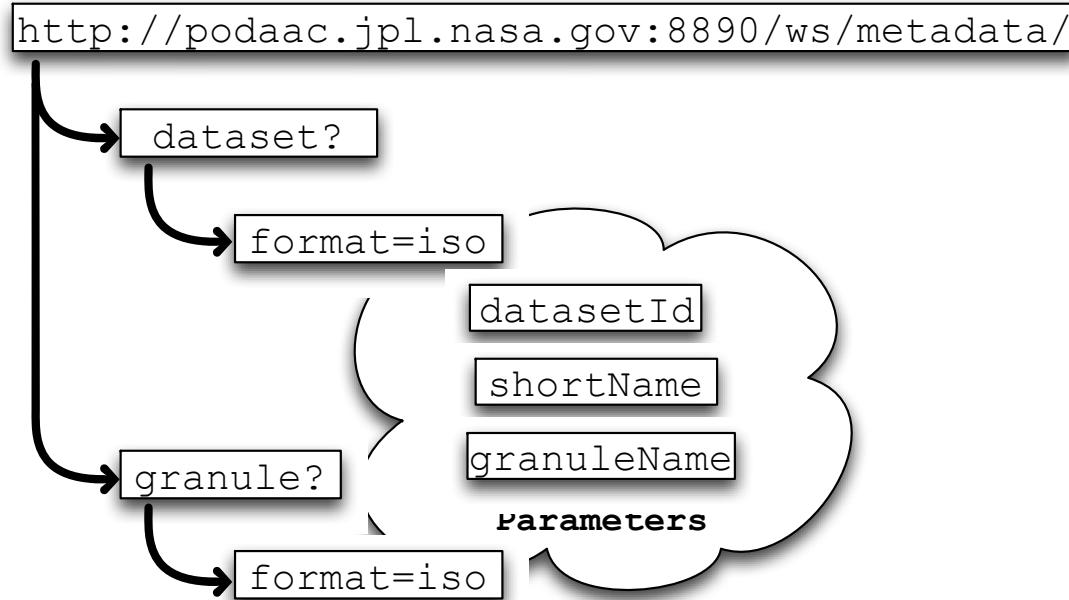


National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



ISO 19115-2



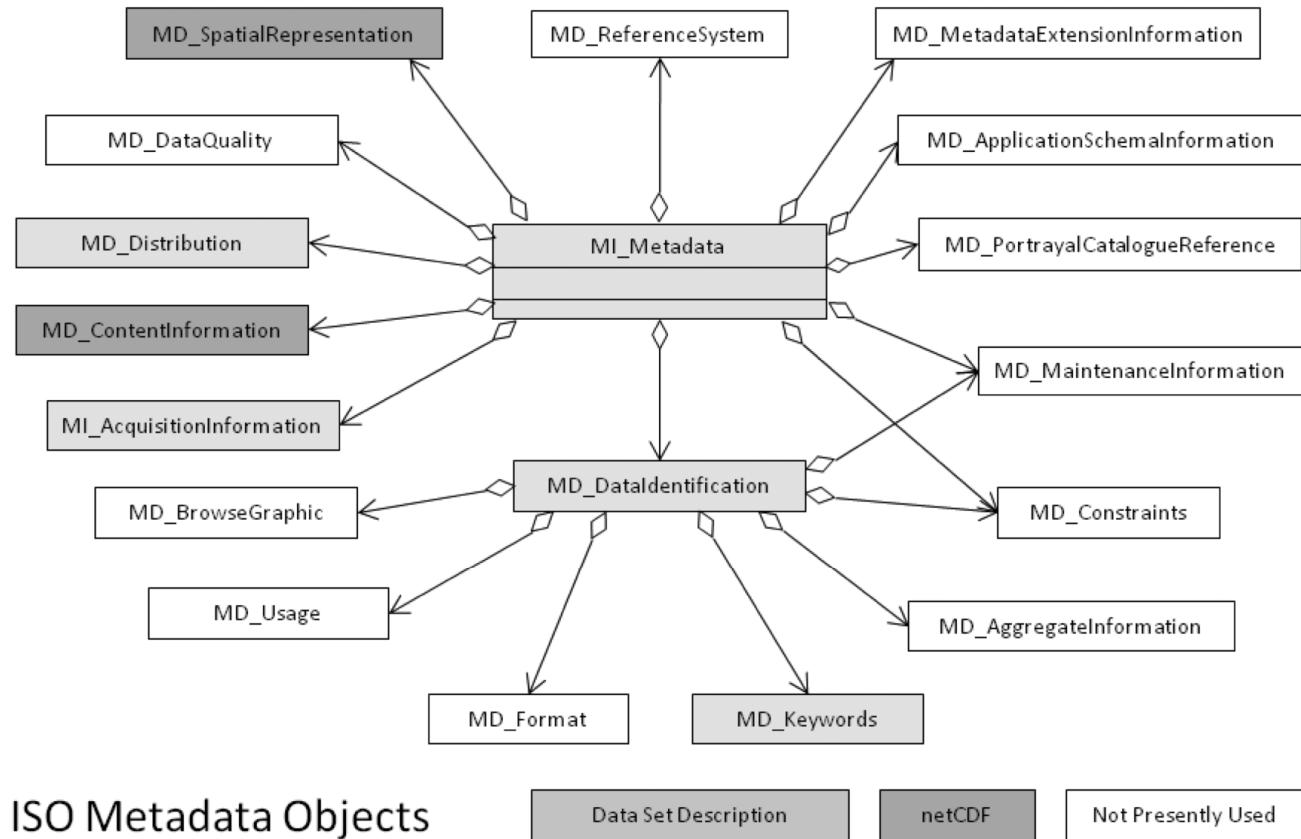
Terminal

```
% curl -X GET \
? "http://podaac.jpl.nasa.gov:8890/ws/metadata/dataset?format=iso&shortName=OSDPD-L2P-MSG02"
```



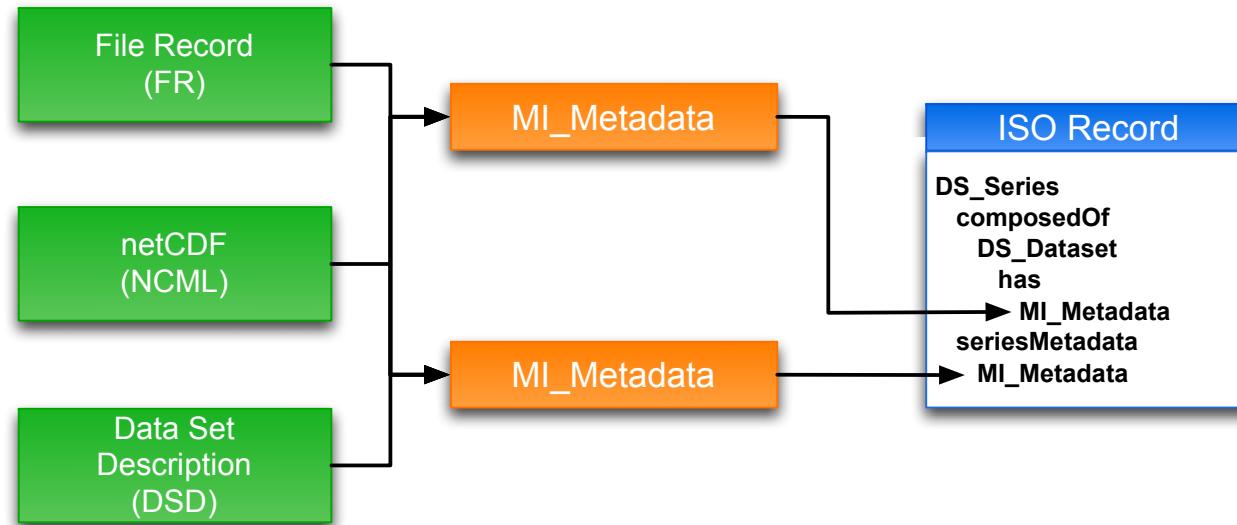
ISO Metadata Model

- Where do we stand?





METADATA TO ISO



GHRSSST Data Processing Specification version 2 on Metadata Conventions depicting the workflow of metadata translation for both data set and granule (file) level metadata to ISO 19115-2.



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



ISO 19115-2 Mapping Challenges

- Certain ISO metadata objects require the following
 - Opening granule file to retrieve necessary information
 - MD_SpatialRepresentation needs dimension size, resolution
 - MD_ContentInformation, specifically, MI_CoverageDescription, needs physical measurement variables
 - Input from provider and/or data engineers
 - DQ_DataQuality needs identification of what the quality flags are per dataset
 - Collection of external information
 - MD_DistributionInfo, for example, needs information about remote distributors, e.g. URL, contact person



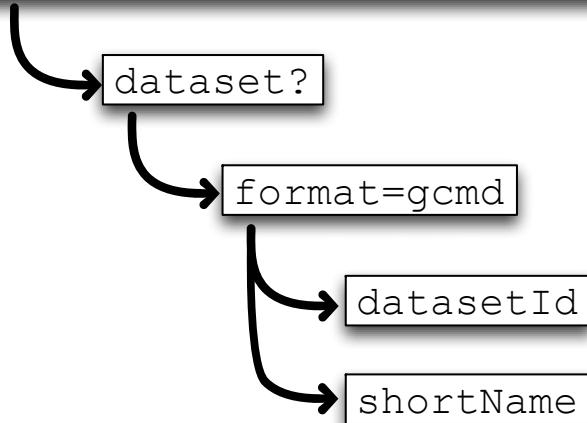
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



Global Change Master Directory

```
http://podaac.jpl.nasa.gov:8890/ws/metadata/
```



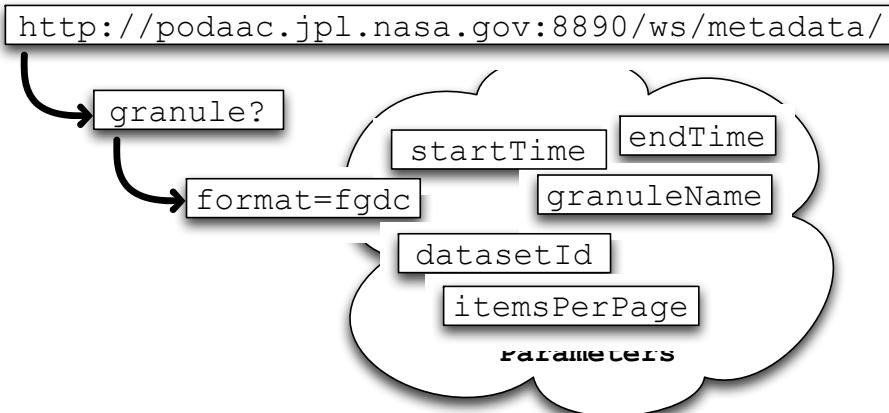
- Moved away from the legacy Excel spreadsheet translation
- DIF records export to NASA Global Change Master Directory

Terminal

```
% curl -X GET \
? "http://podaac.jpl.nasa.gov:8890/ws/metadata/dataset?format=gcmd&shortName=OSDPD-L2P-MSG02"
```



Federal Geographic Data Committee



- Speedup of FGDC record generation
 - For 300+ granules, record generation went from taking hours to seconds
- Use of template engine supports reuse
 - Generate FGDC record for all datasets, not just GHSST datasets
- Retrieve FGDC record as needed via web service call
 - No longer require pre-generate FGDC record files

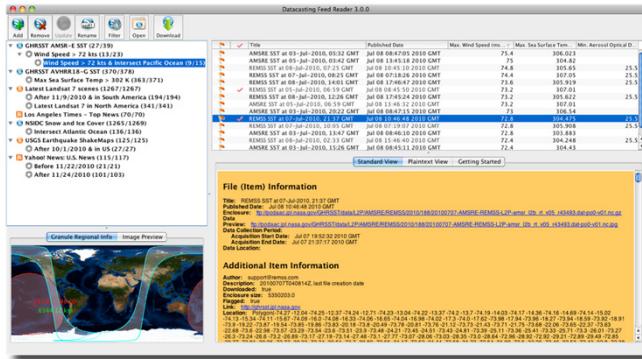
Terminal

```
% curl -X GET \
? "http://podaac.jpl.nasa.gov:8890/ws/metadata/granule?format=fgdc&
shortName=JPL-L2P-MODIS_A&
startTime=2011-02-01T00:00:00Z&endTime=2011-02-01T23:59:59Z&itemsPerPage=500"
```



Future Work

- Metadata quality improvements
 - Data description improvement for consistency
 - Resolving missing attributes (ISO)
 - Consistent dataset Long Names
- Datacasting generation



The screenshot shows the Datacasting Feed Reader 3.0.0 interface. The main window displays a list of datasets with columns for Title, Published Date, Max. Sea Surface Temp., and Min. Aerial Office ID. The datasets listed include various SST and AMSRE products from different sensors (TIR, AVHRR, MODIS, etc.) at different times (e.g., 03/07/2010, 07/21/2010). Below this is a 'File (Item) Information' panel with tabs for 'Standard View', 'PlainText View', and 'Getting Started'. This panel contains detailed metadata such as Title, Published Date, Author, Description, Download URL, File Size, Line, and Location.

- ECHO submission record generation



Summary

- There will always be
 - Metadata standards
 - Different (may be better) ways to look for data
- Why PO.DAAC decides to embrace OCSI
 - No need to redo the plumbing for each new standard
 - Allows us to focus on the domain – metadata standard and metadata resources
- Users expect more intelligent search results
 - Semantic Web has a lot to offer
- Our responsibilities
 - flexible
 - adaptable
 - scalable
 - fast

Simple to Use
Expectation from our Users



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

